

Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future

Article (Published Version)

Manville, Catriona, Morgan Jones, Molly and Chataway, Joanna (2017) Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future. *Journal of Technology Transfer*. pp. 1-27. ISSN 0892-9912

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/69767/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future

**Molly Morgan Jones, Catriona Manville
& Joanna Chataway**

The Journal of Technology Transfer

ISSN 0892-9912


J Technol Transf

DOI 10.1007/s10961-017-9608-6



Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.

Learning from the UK's research impact assessment exercise: a case study of a retrospective impact assessment exercise and questions for the future

Molly Morgan Jones¹  · Catriona Manville¹ ·
Joanna Chataway²

© The Author(s) 2017. This article is an open access publication

Abstract National governments spend significant amounts of money supporting public research. However, in an era where the international economic climate has led to budget cuts, policymakers increasingly are looking to justify the returns from public investments, including in science and innovation. The so-called ‘impact agenda’ which has emerged in many countries around the world is part of this response; an attempt to understand and articulate for the public what benefits arise from the research that is funded. The United Kingdom is the most progressed in implementing this agenda and in 2014 the national research assessment exercise, the Research Excellence Framework, for the first time included the assessment of research impact as a component. For the first time within a dual funding system, funding would be awarded not only on the basis of the academic quality of research, but also on the wider impacts of that research. In this paper we outline the context and approach taken by the UK government, along with some of the core challenges that exist in implementing such an exercise. We then synthesise, together for the first time, the results of the only two national evaluations of the exercise and offer reflections for future exercises both in the UK and internationally.

Keywords Research impact · Research policy · Research Excellence Framework · Science of science

JEL Classification I2 · I23 · I28 · O3 · O32 · O38

✉ Joanna Chataway
J.C.Chataway@sussex.ac.uk

Molly Morgan Jones
mmjones@rand.org

¹ RAND Europe, Westbrook Centre, Cambridge CB4 1YG, UK

² Science Policy Research Unit, University of Sussex, Falmer, Brighton BN1 9RH, UK

1 Introduction

In 2015, the United Kingdom (UK) invested £31.6 bn (€43.5 bn)¹ in research; approximately £20.9 bn (€28.7 bn) came from the private sector and £10.1 bn (€13.9 bn) came from public sources of funding and, of this, £8 bn (€11 bn) was spent by the higher education sector on research (ONS 2017). These are significant investments and are increasingly justified on the grounds that, at a macro level, spending money on research can lead to improvements in societal well-being and through the process of innovation, enhance productivity and efficiency. However, the exact nature of these impacts and benefits that we receive from investment in, and collective generation of, new knowledge are much less clear. In the current economic climate, in particular, the question remains—what does society get in return for this investment? Does this research improve the health and wellbeing of society? What is the nature or size of these returns, benefits, or, more broadly, *impacts*? Thus, the widely deployed justification of spending of research with reference to ‘real-world’ benefits has been accompanied by an attempt to better understand both how that translation occurs and how good performance in relation to non-academic impact can be measured.

In this paper we will discuss this trend and its implications for research and policy-making communities. We begin by reviewing the literature around research impact evaluation in the context of research policy and will draw specifically on the UK’s recent experience in re-designing its national research assessment exercise, the Research Excellence Framework (or ‘REF’) to incorporate and reward the non-academic impacts of research. The impact component of the 2014 REF has been much discussed and, alternately, lamented and celebrated in academic and policy circles alike. The Stern Review, led by Sir Nicholas Stern, has recently published recommendations about future improvements and in doing so confirmed that the measurement of impact will be part of the future landscape of research assessment in the UK (BEIS 2016). In light of that this article systematically addresses the exercise as a whole and considers lessons for the future.

The authors have been involved in the only two primary evaluations of the impact element of the REF which have been undertaken at a national level. We use this experience to draw out lessons and implications and, for the first time, bring the lessons from the two evaluations together with other impact evaluation work in both policy and academic circles. Our studies were designed to understand the process of including the assessment of impact in REF 2014, from the submission process through to the panel’s evaluation of the impact case studies. Our findings and insights are drawn primarily from these two commissioned policy research and empirical evidence gathered over the course of these two studies. The nature of research imposed some limitations on scope and the nature of enquiry, which are discussed below, but also provided exceptional access to policymakers, research stakeholders, leaders of Higher Education Institutes (HEIs) and researchers across the UK.

After highlighting the findings of the evaluation, we will explore the implications for longer term cultural change within the research sector in relation to the assessment of research impact. This will lead us to a series of reflections which highlight the potential future opportunities and challenges for research impact assessments in relation to broader changes in the global research environment.

¹ Conversions done using the average exchange rate for 2015, which was 1.377 Euro to 1 GBP.

2 Background and context for the assessment of research impact

We will start our review of the background and context for the assessment of research impact by first considering how impact has come to be defined and the rationale for why it should be measured as part of a national research assessment exercise. In doing so, we will not attempt to undertake an exhaustive analysis of the origins of the concept of impact, but we will highlight a few historical steps which have been central to its usage today. We draw on both academic and policy literatures as we think is a useful exercise because there has been a considerable body of commissioned policy literature generated in relation to research impact assessment and this article provides an opportunity to bring together and reflect on that policy-oriented research alongside academic literatures.

2.1 The context for assessing impact

In the lead-up to REF 2014, many policymakers argued that the so-called ‘impact agenda’ was not new. Indeed, Hill, a Higher Education Funding Council for England (HEFCE) policymaker who was deeply involved in the development and delivery of REF 2014, wrote a review arguing this point (Hill 2016). Hill traces some of the first codifications of the idea that publicly funded research can, and should, have a practical usage for society to the words of Vannevar Bush in his treaty, *Science The Endless Frontier*:

There are areas of science in which the public interest is acute but which are likely to be cultivated inadequately if left without more support than will come from private sources. These areas - such as research on military problems, agriculture, housing, public health, certain medical research, and research involving expensive capital facilities beyond the capacity of private institutions - should be advanced by active Government support... As long as [universities] are vigorous and healthy and their scientists are free to pursue the truth wherever it may lead, there will be a flow of new scientific knowledge to those who can apply it to practical problems in Government, in industry, or elsewhere. (Bush 1945, p. 4)

Hill goes on to point out these ideas that investment in knowledge can deliver benefits to society proved enduring. Increasingly ideas of a purely linear and spontaneous relationship were supplemented with further thinking about how best to achieve links between academia and broader society, but Bush's basic contention that science was essential to furthering society's social and economic ambitions provided a long lasting rationale for investment in science. In the UK, they can be seen explicitly noted in the government's White Paper, *Realising our Potential* (Cabinet Office 1993), and subsequent investment framework for science and innovation which laid out how science and innovation could contribute to economic growth and the attributes of a research system which would be needed to achieve this (HM Treasury 2004). A wide variety of policy documents from a variety of political and policy angles increasingly focused on maximising the contributions of research and science. As part of the evolution of ideas in this policy area, from the mid-2000s there has been a growing interest in the idea of research ‘impact’ and how it might be measured and encouraged.

One of the first steps in efforts to further understand and measure impact is to recognise that there are many different definitions of impact, as the examples below show (emphases are the authors own).

The social, economic, environmental and/or cultural **benefit** of research **to end users** in the wider community regionally, nationally and/or internationally (Department of Education, Science and Training 2006, p. 21)

‘Pathways to impact’ statements required by the Research Councils include: “Academic impact: The **demonstrable contribution** that excellent research makes to academic advances, across and within disciplines, including significant advances in understanding, methods, theory and application” and “Economic and societal impacts: The **demonstrable contribution** that excellent research makes **to society and the economy**. Economic and societal impacts embrace all the extremely diverse ways in which research-related knowledge and skills benefit individuals, organisations and nations...” (RCUK 2014)

Impact is defined as an **effect on, change or benefit to** the economy, society, culture, public policy or services, health, the environment or quality of life, beyond academia (HEFCE 2011a b, p. 48)

“‘Broader impacts’ are defined as the potential to **benefit** society and **contribute** to the achievement of specific, desired societal outcomes.” (NSF 2007)

Interactions between researchers and society which are ‘**productive**’, meaning that they lead to efforts by the stakeholders to apply research findings and hence change their behaviour in some way. (Spaapen et al. 2013, p.4)

The highlighted words in the various definitions above point out a central and common idea that what is being sought is a diverse set of tangible benefits for society and the economy. However, this is not to say that there is, necessarily, agreement on the philosophical underpinnings, nor definition of impact (Penfield et al. 2014; Greenhalgh et al. 2016). While we do not intend to resolve that debate here, we do aim to point out that there is a common thread running through different definitions, and there are a corresponding set of methods and frameworks that have evolved to try and help researchers and policymakers understand impact.

In many cases, impact has been presented through case studies and the associated benefits arising highlighted using narrative stories and examples.² However, there are many ways to demonstrate the impact of research, and more traditional methods range from analysing the economic returns from research (Buxton et al. 2004; Glover et al. 2014; HERG et al. 2008), conducting sophisticated bibliometric assessments to trace knowledge outputs and flows (Martin and Irvine 1983; King 1987; and May 1997),³ and methods which look at the nature of knowledge interactions and the relative productivity which emerges (Spaapen and van Drooge 2011; Molas-Gallart and Tang 2011). With the growth of the field of research impact evaluation in recent years, newer methods for assessing impact are now beginning to also emerge, including forms of data mining and analysis of multiple, alternative metrics, for understanding how and by whom research is being used (Priem et al. 2012). Each of these approaches and methods can be designed and used to highlight particular societal impacts (Guthrie et al. 2013; Greenhalgh et al. 2016). However, under any circumstances, demonstrating, measuring, evaluating and understanding

² See for example Research Councils UK (2012), as well as case study based frameworks such as the Payback Framework, described in Buxton and Hanney (1996).

³ The literature on bibliometrics is vast. These are just a few early examples of work using bibliometrics and other indicators to track scientific progress.

impact is not straightforward and there are different methodological drivers and corresponding challenges of research impact assessments. It is to this issue we turn to next.

2.2 The challenges of research impact assessment

In addition to the drivers of research impact assessment, there are a number of challenges that need either to be acknowledged or, where possible, addressed in the development of a system or approach to measuring research impact. These challenges have been identified in multiple works over the years (see for example, AMS 2006; ESF 2012), and can be considered to be of differential importance depending on the primary purpose of the assessment exercise itself. Though not an exhaustive list, some of the primary challenges are recognised to include⁴: the time lags between research occurring and the resulting benefits to society, the non-linearity of the research to impact process and the inherent reflexivity of this process, the challenge of understanding attribution versus contribution, and the diversity of impacts which can arise and the tension with impact metrology. We explore each briefly, in turn.

Time lags The time it takes for medical research to translate from ‘bench to bedside’ has been estimated to be on average 17 years (Slote Morris et al. 2011; Hanney et al. 2015). We know that advances in other disciplines, like engineering or physics, may take 50 years or longer to materialise (Illinois Institute of Technology 1968). We also know, though, that impacts can happen before we fully understand the mechanisms which led to them: steam engines were around before we understood the laws of thermodynamics and the Wright brothers flew airplanes before we understood aerodynamics (Kline and Rosenberg 1986). This means that any assessment of contemporary impact may have to look at research that could have occurred two decades ago or, in the context of anticipated impact, two decades hence.

The non-linearity of impact and the engagement with research users Just as we cannot always anticipate the time taken to realise impacts, we must be attune to impacts which may not arise in a temporal, linear manner. For example, Weiss provided us with seven different models through which research translates into policy (Weiss 1979), Walter et al. (2004) have outlined three models in which research is applied to social care practice, as well as five conceptual frameworks in which these models operate, and Boaz et al. (2009) have reviewed the literature in the environmental field of how research has an impact on policy outcomes. At the risk of over-simplification, these literatures and other observations (see for example Hill 2016 and our own evaluation findings below), and insights from the science and technology studies and innovation systems literature (Lundvall 1992; Freeman 2008), confirm that the links between research and impact are often not linear. The longer and more complex the time horizon and pathway to impact, the harder it will be to assess the route to impact.

Related, any routes to impact involve research users and it is acknowledged that they need to be included in any impact assessment exercise, no matter the methodology. A major challenge going into the REF was the question of whether research users would engage with the process. Involvement of stakeholders in providing evidence of impact is essential, and there is some evidence (from our evaluations in particular), that such engagement results in positive benefits for the stakeholder and the researcher in increasing impact awareness. However, if you are outside the academic system, where peer review is a more commonly accepted modus operandi, there are questions about the incentives for

⁴ For a full discussion of these points, see Morgan Jones and Grant (2013).

engagement. The pilot of the REF indicated that altruism was probably the best explanation for engagement in the exercise (Technopolis 2010); that is, users of the research have an interest in seeing it continue. However, there are still questions about how much user engagement is optimal, or needed. The impact trial in Australia had around 70% of its panel composition as research users, whereas the REF, taking into account the sub-panel members who were research users, had around 27% of its panel composition as research users (Morgan Jones et al. 2013).

Attribution versus contribution This non-linearity relates directly to the challenges inherent in defining not just the contribution of research to impact, but the attribution as well. At its simplest, a research project is likely to have multiple research inputs—different research grants, collaborators working in different research institutions, researchers at different stages of a research career—and result in multiple research papers (outputs). These multiple linkages get increasingly complex as one progresses downstream to assess research impact. The challenge of any system that assesses research impact is to ensure that we have an understanding of the ‘contribution’ and ‘attribution’ relative to the outputs, outcomes or impacts that result from the research input and activity. Here, contribution refers to the relative efforts made by a research team(s) and the relationship to the outputs, outcomes and impacts, whereby attribution refers to the proportional extent to which the outcomes or impacts have resulted from those efforts and research outputs. The way contribution and attribution are, or should be, highlighted will be differentially important depending on the purpose of the assessment.

Diversity of impacts and impact metrology With the increasing focus and study of impact comes the recognition that not only are there a range of impacts which can be identified from academic research, but that these impacts can be captured and measured in different ways. HEFCE identifies eight overarching categories of impact—the economy, society, culture, public policy, public services, the environment, quality of life, and health (HEFCE 2012)—while the synthesis of impacts identified in the REF conducted by the Policy Institute at King’s and Digital Science, London, identified 60 impact topics across the 36 Units of Assessment (KCL and Digital Science 2015).

However, with such a diversity of impacts come challenges in how we measure them. While some areas of impact lend themselves to readily available metrics, such as counts of patents or numbers of patients reached, even these figures can be open to interpretation. Ovseiko and colleagues analysed the submission to the REF pilot of the University of Oxford’s clinical medicine department and systematically analysed many of the indicators suggested in the REF pilot in order to explore their strengths and weaknesses (Ovseiko et al. 2012). Their analysis demonstrates that numerical indicators that appear to be relatively straightforward indicators of wider impact have significant limitations. For example, counting numbers of patents says nothing about their quality, nor whether those patents went on to generate any income. In another example, they point out that figures about returns on investment from intellectual property can also be misleading. Oxford’s innovation spin-out group reported £9.8 m in returns from 2004/05 to 2008/09, 50% of which could roughly be attributed to clinical medicine. However, when compared against the inputs to the department of around £612.9 m over the same period, the actual return was only 0.8% and therefore not as impressive as the initial crude total might have suggested.

These examples draw our attention to the limitations of using any single indicator to demonstrate impact; highlighting only one element risks distorting impact and hiding a wider picture. There are challenges, then, in developing sets of impact indicators which could ‘close down’ wider analyses that are necessary to appreciate the full context.

The resolution of any of these challenges is inevitably linked to the purpose and objectives of the assessment exercise, for instance an exercise done for allocation will need to ensure there is an equivalent starting point across the system, whereas an assessment for advocacy purposes may be less concerned with this point.

2.3 The use of impact in the Research Excellence Framework

The UK has been one of the first countries to formally include impact as a metric in its national research assessment exercise, the REF. The REF, and before it the Research Assessment Exercise (or RAE), is one of the main vehicles for the allocation of over £1bn research funding annually to universities in the UK from the four Higher Education Funding Bodies and assessment has occurred on a (near) quinquennial basis since 1986. The RAE assessed research excellence in universities by the quality of research outputs and other measures of the research environment, including research students, income and evidence of esteem (HEFCE 2008).

Overall, the REF assesses universities at a unit or discipline level on three primary metrics: the basis of the *quality* of research outputs, the *vitality* of the research environment and the *wider impact* of research. We won't focus here on the first of these two objectives of the REF, except to say that they have, more or less, remained relatively unchanged in their aims since the concept of a national assessment exercise was first introduced in 1986 (Hicks 2012).

After initial reviews commissioned by HEFCE (Grant et al. 2010), a pilot exercise (Technopolis 2010), and an extended consultation period, HEFCE issued its REF guidance in two parts. First, in 2011 the Assessment Framework and Guidance on Submissions was published (HEFCE 2011a b), followed by the Panel Specific guidance documents in 2012 (HEFCE 2012). The approach favoured by HEFCE involved the use of case studies. The assessment of impact in the REF was based on expert review of case studies submitted by HEIs in the UK, using an approach that largely follows recommendations from an international review of practice from 2009 (Grant et al. 2010). Impact was assessed in the REF through case studies.

Case studies describe specific impacts that occurred in the assessment period; 2008–2013 for REF 2014. Case studies are presented as a four page summary with sections for: Title, Summary of impact, underpinning research and references, details of impact and sources of corroborating evidence. Taking into account time lags, research could have been conducted in the last 20 years. There were a number of eligibility criteria, such as a 2-star quality for the underpinning research, and the research musty have been conducted at the institution claiming the impact. The number of case studies was dependent on the number of staff submitted, with one case study required per ten members of staff, with a minimum of two case studies for each submission.

UK HEIs submitted 6975 case studies across 36 units of assessment. The assessment was conducted by an elected panel of academic experts and eminent research users. Impact was graded on a 4-star scale ranging from modest to outstanding impact using the criteria of 'reach' and 'significance'. Initially the panel members underwent training and a calibration exercise to align their understanding and measurement of impact. Then each case study was reviewed by at least two assessors—at least one academic panel member and one research user. These individual scores were then discussed and moderated and a score allocated. Moderation was also conducted at a sub- and main panel level. Where there was uncertainty around the credibility of the statements being claimed, assessors could raise an audit query and the facts were validated, through evidence.

As has been pointed out elsewhere (Hill 2016), the REF was a new and unprecedented attempt at incorporating impact into a research assessment exercise whose primary purpose is to allocate research funding. In doing this, the UK was, to some extent, conducting an experiment as to whether universities could demonstrate their impact, and how this would be done. This latter part is where our focus lies; there was little question in many people's minds that research funded by the government was having an impact, but there were many questions as to whether it could be demonstrated at scale and across multiple disciplines, and evaluated appropriately. There are thus two main questions: did the process work, both in preparing the impact submissions and assessing them; and what kinds of impacts did we see?

We will spend the next part of this paper discussing our methodology and findings in order to answer the first of these questions. The second question is, to a large extent, a work in progress and we reflect on this and the implications for the future of exercises like the REF in the discussion section below.

3 Methodology

We were involved in the two studies to address the first question about the process of the impact assessment (Manville et al. 2015a, b, c). We conducted two evaluations in order to assess the *process* of evaluating impact by first evaluating how universities *prepared* their impact submissions and then how the panels *assessed* impact. Our studies were designed to understand the process of including the assessment of impact through REF 2014, from the submission process through to the panel's evaluation of the impact case studies. The aim across both evaluations was to explore the strengths and weaknesses of the process, assess the consequences and implications, and provide recommendations for improvements for future assessments. The research used a mixed-method approach, including document review, interviews, focus groups, online surveys and cost analysis. Across the two parts of the study—submission and assessment we came up with a number of key findings. The following section provides further details on the methodology used for these studies.

3.1 Methodology to assess the submission process for the impact component of REF 2014

This first evaluation study collected data from three main stakeholder groups: those associated with the leadership and administration of the impact assessment element of REF 2014 in HEIs; those who led the development of impact case studies (i.e. research academics); and research users (i.e. the beneficiaries of research). These stakeholder groups were comprised primarily of individuals from a sample of 18 HEIs in England were randomly selected from the population of 123 HEIs who indicated their 'intention to submit' to REF 2014. HEIs were selected to oversample institutions making larger submissions, but also to ensure representation of the smaller ones at the same time. These 18 institutions were supplemented by one HEI in Wales and two HEIs in Scotland, each chosen by their respective HE funding councils.⁵

There were five main data collection and analysis methods used in the research: site visits to 21 institutions selected for our sample; a survey of academics and others involved

⁵ For details of the individual institutions and a full discussion of the sampling strategy, see Manville et al. (2015b).

in the impact preparation process at those 21 institutions; a research user interview; a cost-benefit analysis; and triangulation and analysis across the data collected.⁶

Site visits We conducted site visits in order to gain qualitative insights into the process that HEIs went through in preparing submissions and to understand the benefits, challenges and consequences they perceived. We conducted two site visits to each of the 21 HEIs in our sample, with the first visit simply introducing the study and the second site visit being the main vehicle for data collection. The main site visits occurred between December 2013 and February 2014 and during these visits two researchers from the evaluation team spent from half to one day at the institution (depending on the size of the institution) conducting semi-structured interviews and small focus groups with a range of individuals involved in leading and supporting the impact submission for the HEI. The individuals included senior leaders (Vice Chancellors, Pro-Vice Chancellors, Department Heads, etc.), senior administrators (REF Coordinators), UOA leads within departments, and impact support officers. In total, we held 126 interviews and met with 327 individuals during these second site visits.

Detailed notes were taken during each interview and were then coded using QSR NVivo 10 International software. The research team developed a site visit NVivo code book which included: descriptive nodes for each participating HEI and UOA, generic nodes that covered interviewee type and either positive or negative views about the topics covered, and thematic analytical codes which were based around the interview protocol, for example descriptions of benefits or challenges. In total there were 48 nodes in the code book and each set of interview data from a site visit was coded by a member of that site visit team and reviewed by the second member.

Impact case study lead author survey Impact case study lead author surveys were also conducted in order to ensure that the views of those who worked directly on these documents (and may not have been present during site visits) were captured.⁷ Institutions were asked to identify 'lead' author(s) for the impact case studies within their submission, and these individuals were contacted by the evaluation team to complete the surveys. The surveys focused specifically on the process of producing the impact documents required by the REF. They included two main types of questions: (1) questions about different estimates of the amount of resource (e.g. time and people) required to produce the documents and (2) qualitative questions about the benefits and challenges of the process, notable practices employed, and suggestions for improvement. The survey was open for each HEI for four weeks after the site visit was conducted. Surveys were hosted through SelectSurvey.⁸ As detailed in Table 1, for the impact case study author survey, 1793 individuals were identified by the HEIs in our sample and invited to take part. The response rate across all 21 HEIs ranged from 36 to 92%, with a mean response rate of 54%.⁹

The data were analysed qualitatively and quantitatively where appropriate. The analysis of the descriptive statistics was conducted in Microsoft Excel and involved calculating the

⁶ Full details of the methodology for each element is described in Manville et al (2015b).

⁷ We also conducted a survey of lead authors of the 'impact template' document required by the REF, a more strategic, overarching document which was submitted by each Unit of Assessment. As the analysis of these documents is less central to the focus of this paper, we do not provide detail on this component here.

⁸ SelectSurvey is the online survey tool used by RAND and hosted by the RAND US Information Science and Technology (IST) group. See <http://selectsurvey.net/>

⁹ Due to issues of confidentiality and the need to provide anonymity for survey respondents, we were unable to link respondents to specific UOAs and therefore could not provide any systematic analysis of non-response rates across the sample.

Table 1 Response rate to the impact case study and impact template surveys

Type of survey	Number of individuals invited to complete the survey across 21 institutions	Number of respondents	% of response rate
Impact case study (REF3b)	1793	962	54

median and interquartile ranges for the questions addressed. Due to the breadth, depth and diversity of views reflected across the open text responses, these were qualitatively analysed in a similar manner to the site visits using NVivo (following a code book as described above).

Research user interviews Research user interviews were used to ascertain how research users engaged with REF 2014, and whether the process of providing evidence to researchers produced any particular benefits or challenges. Short, 15–20 min telephone interviews took place using a semi-structured interview protocol. In total, 23 individuals and six organisational representatives were interviewed. The sample was generated from a list provided to us by HEFCE of individuals from the 21 institutions in our evaluation who were cited as either contactable for corroborating an impact case study, or who had provided testimonials in support of an impact case study.

Cost analysis In order to estimate the costs of preparing submissions for the impact assessment element of REF 2014, all HEIs were asked to complete a cost estimation worksheet. This worksheet asked for estimates of costs and other resources relating to the preparation of the impact component of the REF submission in three main categories: type of costs, type of activity involved in the preparation, and proportion of costs estimated to be ‘start-up’ costs. From this data we were able to generate four key indicators: median cost per impact case study; median cost per impact template; total costs of the impact portion of REF 2014; and transaction costs (i.e. total costs divided by estimated QR benefit/funding). We also examined, where possible, the cost drivers (i.e. activities undertaken in developing the impact case studies or impact templates), and differences by HEI characteristics (i.e. size of submission).

3.2 Methodology to assess the evaluation process

This second evaluation study collected data from individuals who assessed the impact assessment element of REF 2014 and worked on the panels. There were five main data collection and analysis methods used in the research: documents review; focus groups; a survey of those involved in the reviewing impact case studies; Interviews with individual panellists; analysis of scoring data; and triangulation and analysis across the data collected.¹⁰

Document review The research team conducted a document review of publically available material that is linked to the REF 2014 process, paying particular attention to details of the assessment process. All documents were reviewed prior to undertaking fieldwork to develop an understanding of the process, and inform the protocol development. Documents included the results of the impact pilot (Technopolis 2010), guidance on

¹⁰ Full details of the methodology for each element is described in Manville et al. (2015c).

submission (HEFCE 2012), and main panels' reflections on the REF process (HEFCE 2015) and documents were made available to the research team via HEFCE.

Focus groups Two focus group meetings of panel members were convened by HEFCE in late November 2014. The first was for research user panel members, and the second was for academic panel members. The session on the impact element lasted 1 h and discussion was facilitated by two members of the team from RAND Europe. The focus groups varied in size, from 9 to 20 panellists, with a total of 132 individuals attending. Due to this size and limited time, we conducted a survey in advance of sessions asking a range of questions on perceptions of the process, receiving 79 responses. The themes for discussion in the focus group were selected based on areas of disagreement or consensus from these surveys, in particular: scoring, moderating and calibrating impact case study assessments; working with the REF's definitions, rules and templates; and the use of corroborating evidence in informing impact case study assessment. Data were recorded in notes and then coded using QSR NVivo 10 International software. The research team developed an NVivo code book for the analysis along the lines described above for the site visit data in the first evaluation.

Interviews with individual panellists The aim of the interviews was to understand the process of assessment and panellists' perceptions of the process. Twenty interviews were conducted, and these were divided between panel advisors, sub-panel members and impact assessors. The protocol was framed around the different parts of the impact case study assessment process, providing context and detailed understanding on the effectiveness and suitability of the rules and guidance, the training process and the assessment process itself. Interviews also explored unforeseen issues and resolution processes used by each panel. Interviews were conducted on the phone and were recorded for note taking purposes.

Survey of all individuals involved in the assessment of impact in REF 2014 A survey was sent to all 1161 panellists involved in the impact element of the assessment. The purpose was to ensure that the views of all of those involved in the process were captured. Panel members were classified as sub-panel impact assessors, main panel users, main panel members, advisors, secretaries and academic sub-panel members. The surveys were piloted at the end of October and modified based on feedback. The main survey was then open for four and a half weeks in November and December 2014. Respondents were sent a personal link and then two reminders, one halfway through the survey window and one on the closing date. Response rates varied from 47 to 69%, by type of panel member with an overall response rate of over 49% across the sample. The data was analysed in Microsoft Excel.

Analysis of the scoring data We were provided with two datasets by HEFCE—the scores for each impact case study and impact template submitted to a specific sub-panel (UOA)—as well as the overall submission results by institution, at the sub-panel (UOA) level. These datasets were then interrogated to further understand the process which took place—for example the allocation of case studies, the consequences of the levels awarded and the relationship between impact case study and impact template scores.

3.3 Limitations of our methodology

Our studies aimed to minimise the limitations of the methodology by triangulating different data sources and methods. However, it is important to acknowledge the limitations and caveats of our study. When conducting fieldwork, through site visits, interviews or focus groups a semi-structured interview protocol was used which meant not all questions were asked at all interviews. In addition, contradictory points could be raised within one subset of our sample, i.e. within an HEI or unit of assessment, and therefore the range of

views provided is presented. There may also be response bias in the survey, for example people who were either overly positive or overly negative about the process may have been more inclined to respond. Finally, our cost estimations relied on collecting time and cost estimates through surveys. The data was self-reported and reported retrospectively, so the accuracy may vary.

4 Findings: What have we learned from research impact assessment exercises?

4.1 Overview

In this section we will present empirical evidence from our work evaluating the REF, to illustrate three main findings: there were benefits and burdens associated with the assessment of research impact as part of the REF; there are issues associated with articulating research impact which presented different challenges and opportunities for the sector; and there are a series of implications for the sector to grapple with going forward.

4.2 Benefits and burdens of assessing research impact as part of the REF

Our evaluation, and other analysis commissioned by HEFCE, showed that there were significant benefits, as well as burdens, for the sector as a result of assessing the impact of HEI research as part of the REF. We highlight the following main areas of benefit and burden below: the fact HEIs could, and did, articulate impact; cultural change; cost burdens; and intangible burdens.

The articulation of impact One of the most obvious benefits we identified through our evaluation was the simple fact that impact could be articulated. In our survey of those involved in the REF at our sample of HEIs, 48% of respondents across all 21 HEIs in the sample indicated that identifying and understanding impact was one of the three main benefits of the REF, making it the most frequently cited benefit amongst respondents.¹¹ Given that the REF was the first national assessment exercise of its kind to attempt to systematically undertake a retrospective impact assessment, the fact that 6975 impact case studies were submitted shows that impact could be identified, understood and articulated. This should not be overlooked.

Further support for this comes from subsequent analysis of the case studies themselves. An analysis commissioned by HEFCE showed the diversity, and global nature, of the societal impact of research from UK HEIs. From the 7 main areas of impact described in the definition, Kings College London and Digital Science used topic modelling techniques to identify 60 impact topics, ranging from animal husbandry and welfare to informing government policy to work, labour and employment. The study also found that impact case studies were based on research that was interdisciplinary in nature, with 78% of case studies citing research from 2 fields or more, and presented impacts in multiple areas (KCL and Digital Science 2015).

Cultural change and appreciation of impact In addition to the identification of impact, those who participated in our evaluation study also highlighted that there was a value and benefit to understanding impact. This was seen in the survey findings noted above, and

¹¹ As noted in the methodology section above, this was an open text box question, responses to which were then coded using the qualitative analysis software, NVivo, in the analysis.

confirmed through the site visits we conducted. One of the most frequently cited benefits from across the site visits was that the process allowed researchers to comprehend the impact of their work to a greater extent than before the exercise and even allowed researchers to identify previously unknown impact arising from their research.

[It] highlight[ed] the broader way in which my research had impacted on society, sometimes quite unexpectedly (Manville et al. 2015a, p. 8, statement from impact case study survey respondent)

Other benefits identified in our survey included such as: increased recognition for individuals within HEIs of those undertaking impact activities (33% of total respondents to the case study survey); the stimulation of broader strategic thinking about impact (66% of total respondents to the impact template survey and a major theme identified from the site visits); and the review and reaffirmation of relationships with external stakeholders (22% of total respondents to the case study survey and a major theme identified from the site visits). It is worth noting that three of these, affirming relationships, understanding impact, and recognition were also benefits that were identified in RAND Europe's evaluation of the Excellence in Innovation for Australia (EIA) trial, suggesting that there are some similarities across research impact exercises which may be beginning to emerge (Morgan Jones et al. 2013).

The cost burdens for the sector Alongside these benefits were also burdens associated with the process. In our evaluation, we estimated that it cost universities £55 m (with a range of £51 m–£63 m) to prepare impact submissions as part of REF 2014 (Manville et al. 2015a). This broke down at an estimated cost of £7500 per impact case study produced and £4500 for an impact template describing the unit's strategy to facilitating impact, and an estimated time estimate ranging from 8 to 30 days to produce a case study.¹² Within this total figure, there was some evidence of economies of scale: for institutions that produced more than 100 case studies, the median cost per case study was under £5000, compared to over £8500 for those that produced less than 100.

A more nuanced look at the data show that training and other educational activities aimed at building understanding of what impact was accounted for nearly one-third of costs reported by the institutions. This interesting in light of the finding above that understanding impact was the most frequently cited benefits of the REF process. It also suggests, though, that this was part of the 'start-up' costs of the REF and may not be such a significant component in future iterations of the exercise.

In terms of assessment, based on the quantitative data collected through the survey, the median amount of time spent on the process by impact assessors was 11 days. There was, however, some variation in the commitment made by individual impact assessors—the interquartile range of the estimates of time spent by impact assessors from the survey spanned from 8 to 16 days. This burden was particularly significant for impact assessors (rather than academic panel members), since though time out of work may have been given to cover the time attending meetings, they indicated they had to find the time for the assessment of impact documents themselves in addition to their full time job (in contrast to academic participants who may have been able to assess case studies and impact templates during working hours).

Combining the cost estimation of the submission and assessment elements of the impact component of REF 2014, we get an estimate for the total burden of the impact element of

¹² As explained in Manville et al. (2015b), for the cost estimates we erred on the higher side of this estimate and used the 30 day estimate.

the assessment process of £53m to £66m (based on the interquartile ranges as described above and the assumptions set out in Manville et al. 2015c), with a best estimate of £57 m. In comparison to the amount of money allocated by the REF, however, the transaction costs of the impact element were only 3.5%, which in comparison to the transaction costs of research grants, which has been estimated closer to 10%, is relatively small (DTZ Consulting and Research 2006).

It is also worth noting that we can compare the costs of the REF with other exercises of a similar nature, notably the REF pilot exercise and the Excellence in Innovation for Australia exercise. In both cases, the time taken to prepare case studies was remarkably similar. For the REF pilot exercise, it was estimated that it took 3.1 staff days to prepare an impact case study, along with 2.1 ‘coordination days per output’ (Technopolis 2010). For the Australian exercise, it was estimated that each case study took 3 days to produce (Morgan Jones et al. 2013). These figures are clearly much less than those estimated for the REF, but neither the pilot exercise nor the Australian exercise had any funding attached to them. Thus, we might conclude that the much greater time estimates for the REF exercise are to do, in part, with the greater formality of the exercise and the monetary implications associated with it. This, we believe, resulted in a significant ‘gold-plating’ effect which resulted in additional amounts of time spent on the case studies by individual officers and institutions.

Intangible burdens There were also a number of intangible, or certainly less easily quantifiable, burdens which emerged as part of the REF. Over the course of our site visits it became clear that in a number of institutions the burden of producing the case studies was concentrated in a relatively few number of staff, primarily those designated as impact case study authors. Our assessment indicated that two thirds of the work on any individual case study was conducted by one person, and in many cases this meant academics involved with preparing the submission had to take a break from research activities. This insight from the site visits was confirmed in the survey, with 25% of case study respondents commenting that time was a major challenge in preparing the case study. This resulted in these staff bearing much of the burden for writing the case study, or overseeing a significant number of case studies within a given department or Unit of Assessment. Again, this was a major theme identified from our site visit analysis and the quotations below are indicative of the responses we had.

During the past year, I have written zero papers, I have not given the usual attention to gaining research funding and I believe that the process... has been disastrous for my research and potential, and potentially my own growing international reputation. (Manville et al. 2015a, p. 9; statement made by a researcher during a site visit)

It subtracted a significant amount of time from more central academic duties, such as research. (Manville et al. 2015b, p. 40, statement made by a respondent to the impact case study survey)

4.3 Articulating impact as part of REF 2014: challenges encountered and opportunities realised

Though the ability to articulate impact was identified as a main benefit of the REF process through our evaluation, the very act of doing so presented a number of challenges and opportunities for the assessment of research impact. These challenges align with those that were discussed earlier, and each is discussed in turn.

Time lags The REF exercise set a specific time window for the admissibility of a case study both in relation to the time during which the impact occurred and the time in which the research took place. The research window (1 January 1993–31 December 2013) set time constraints within which research underpinning the claimed impact must have been conducted and published. The impact window (1 January 2008–31 July 2013) set time constraints within which the impact must have occurred in order to qualify for inclusion. In the case of the research window, some participants in our site visits thought that the 20 year time window was an opportunity to reflect and draw upon research conducted over a long time period, while others felt that the window was too short. There was, though, no majority of opinion observed in either case. As was discussed above, the time lags for research benefits to accrue outside academia and be observed in wider society is known to vary. While the figure of 17 years is generally accepted and cited for the time lags in medical research (Slote Morris et al. 2011), there is less of a literature around other fields, particularly the social sciences, physical sciences and arts and humanities.

What was seen to be a more significant challenge by the majority of participants in our site visits was the impact window. Specifically, some participants in our evaluation felt that the window did not account for the fact that impact was a continuum that could start at any time, before or after the research, and before, during or after the impact window. They were unsure as to how this continuum, and indeed non-linearity (see below) of impact should be reflected in the case studies. Some respondents to the survey, in commenting on how HEFCE could improve the rules, felt that the ‘insistence on impacts limited to the recent past was an artificial constraint’ and that ‘more thought needs to be given to more reasonable inclusion criteria. A 20-year cut-off for time [from research] to impact is completely unreasonable’ (Manville et al. 2015b, p. 46, statement made by a survey respondent).

The non-linearity of impact One of the common criticisms we observed over the course of our evaluation was that the actual document that academics had to submit the case studies on reinforced an overly linear way of thinking about impact. The impact case study template document had five main sections: a summary, a description of the underpinning research, a list of the main academic publications, the description of the impact, and a section on supporting evidence for the impact. The whole document could not exceed four pages.

One can see how the sequential order of the document itself implied that research led to impact in a linear, one-way fashion, rather than the iterative, and often reflexive, manner in which we know it occurs. In fact, one of the most commonly mentioned improvements to the guidance documents which was given in the impact case study survey was to ‘change what was perceived by many to be an overly linear definition of impact to reflect the more iterative relationship between research and impact (particularly in relation to underpinning research)’ (Manville et al. 2015a, p. 29). Moreover, across our site visit interviews, one of the most commonly identified challenges related to the definition of impact was its overly linear nature (Ibid, p. 15). Interviewees felt that it did not reflect the full range of research impact, nor the way in which impact could be achieved. For example, some thought that impact was not always related to a specific output of research, rather the expertise of the individual resulting from a continuous body of work over a career. In some subjects, such as practice-based research, it is also difficult to establish causal chains from research to impact, as the links can be indirect and impact can often occur before the research results are published (Manville et al. 2015b, p. 16). Again, these findings are consistent with what we know from the literature, as highlighted above, and we already know from the recent

Stern Review that this issue will be seriously considered by HEFCE in the next iteration of the REF (BEIS 2016).

Assessing attribution versus contribution This issue is one that the authors, and others studying research impact, have often noted as problematic in the assessment of research impact. As discussed above, many exercises account for *contribution* to a given impact, but no exercises make any effort seriously to consider *attribution*, or proportional efforts by any one party towards an impact. The closest that the REF comes to this in the guidance documents, where it states that the impact case studies must show that the research made a unique and demonstrable contribution to the impacts claimed:

‘Underpinned by’ means that the research made a distinct and material contribution to the impact taking place, such that the impact would not have occurred or would have been significantly reduced without the contribution of that research. Each case study must explain how (through what means) the research led to or contributed to the impact, and include appropriate sources of information external to the HEI to corroborate these claims (HEFCE 2012, p. 16).

However, for us this just indicates that HEFCE was looking for a clear demonstration of the research’s contribution to impact; there was no requirement to estimate proportional effects and this did not come up as a point of discussion or debate by the panels. Interestingly, though, it was raised as an issue by research users in our evaluation, as well as survey respondents. Two organisations out of the six organisations interviewed noted that it was difficult to relate impact to any specific piece of research carried out by a single higher education institution. They both felt in these cases that the challenge of both attribution and contribution was acute and that in some cases institutions overstated the contribution their research had made (Manville et al. 2015b, p. 56). Impact case study survey respondents also flagged the challenge of demonstrating attribution, although it was not one of the top five challenges identified (Ibid p. 39). Nevertheless, as we will discuss below, any assessment exercise which is concerned with the allocation of money should probably undertake to include an assessment of attribution in its analysis.

Evidencing impact and impact metrology In relation to the preparation of the case studies themselves, we found that the requirement to actually *evidence* impact in a way which meant the criteria of reach and significance was one of the most challenging aspects. As set out in the REF guidance documents, ‘Each case study must include evidence appropriate to the type(s) of impact that supports the claims, including who or what has benefitted, been influenced or acted upon’ (HEFCE 2012). This meant that researchers had to provide independent and verifiable evidence of the impacts. This was, together with the development of a shared understanding of impact, the biggest challenge in producing REF case studies, with 68% of all impact case study survey respondents across all institutions in our sample mentioning it as a significant challenge and over 50% of the same group of respondents highlighting the requirement to evidence claims of impact as an either somewhat or very challenging component of the REF guidance. This majority view was also seen in our analysis of the site visit interviews (Manville et al. 2015b p. 21–23).

The difficulties in providing evidence came down to a few main issues: challenges in gathering particular kinds of evidence, challenges in connecting with research users, and a perception that quantitative indicators of impact were preferable to qualitative evidence (and frustration that the former were not always possible). The types of impact that were identified as most difficult to evidence were: policy changes, cultural impacts, international impacts, or impacts on public awareness (Manville et al. 2015a, p. 17). Though the reasons for these areas being particularly difficult varied, a common theme was the difficulty of

providing clear evidence of impact. For example, for cultural impacts such as changes to attitudes, behaviours, or preservation of artefacts, there is often no baseline and data are not routinely collected. This makes demonstrating impact, and providing firm evidence, difficult.

For example, one respondent commented that the HE funding councils 'need to embrace the spirit of diverse impacts and recognise that many are not easy to demonstrate via quantitative evidence'. Another pointed out:

[It] would be useful to have a much clearer and more nuanced policy which recognises the complexity and diversity of impact, with a much greater emphasis on qualitative impact and innovation. (Manville et al. 2015b, p. 46, statement given by a respondent to the impact case study survey)

While few in our evaluations commented specifically on impact metrology, it was clear from the concerns about evidencing impact that the pressure to find quantifiable ways of evidencing ones impact was a concern across the sector. Many felt that qualitative evidence was not sufficient, though this was not found to be the case in the evaluation of the assessment process. However, given the analysis provided by Ovseiko et al. (2012) above, one can appreciate the difficulties that even quantification provides. We comment on this further in the discussion section below.

The lack of detailed records also meant that evidence had to be reconstructed in many cases. Evidencing case studies therefore required reconnecting with the beneficiaries and 'research users', sometimes after many years, and asking them a series of questions about how the research originally conducted had led to benefits that were being experienced today. In addition to this being a perceived burden on research users, it also raised significant questions related to attribution and contribution. To what extent could researchers claim follow-on and additive impacts if their research fed in at some early stage, but was not the only one? These factors all contributed to many researchers echoing sentiments similar to that of this researcher:

Evidence was the most difficult element of creating the [impact case study] document and as a result, you ended up thinking more about the evidence rather than the impact. The question was, can I evidence this? (Manville et al. 2015b, p. 21, statement made by a researcher during a site visit)

This sentiment was echoed in the evaluation of the assessment process. Academic panel members noted challenges around the use of evidence here related particularly to issues around corroboration and the level of information available in the case studies. Looking at the specific responses, there was a sense that the statements had to be taken at 'face value' and panel members typically did not have access to evidence around the claims made which was a challenge in assessing the case studies. Sub-panel members were also concerned about the extent to which they were able to fairly assess different types of impact. One respondent referred to this as 'comparing apples and oranges', reflecting the challenge of judging very different types of impact in an equivalent/fair manner.

4.4 Implications for the sector of assessing research impact

There were several implications for the sector as a result of the assessment of research impact as part of REF 2014 which emerged from our two evaluations. We cover three main ones here: the changes in practice within HEIs in the UK; the potential for the exercise to

affect future research trajectories; and the representativeness of the impacts presented and the implications this perception will have for the future.

Changes in practice and culture in the sector We found strong evidence that the assessment of impact as part of REF 2014, along with other policies such as RCUK's 'pathways to impact' statements, has led to cultural change within HEIs in our sample, at an individual and institutional level (Manville et al. 2015a). At an institutional level, even a few months after the deadline for REF 2014 (in early 2014) new practices were observed such as including impact as a criterion for promotion; creating institutional or departmental strategies; building a plan for impact into research studies; installing systems to capture impact in real time; using impact case studies for marketing and advocacy promoting their research and value at a regional level; and using impact to support student recruitment. At an individual level, some academics and central university staff commented that it gave them a greater appreciation of the work their colleagues do outside academia, rewarding those who had been engaging with wider society. It boosted the self-esteem and morale of researchers involved in these areas and improved their parity of esteem with 'pure' researchers. One interviewee remarked that preparing for impact assessment 'shone a light on the underplayed and undervalued' and others reported changes to staff promotion and reward schemes to recognise contributions to the process.

Implications for future research trajectories However, linked to changes in practice, it is worth reflecting on the potential effect of the measurement of impact on shifting the research agenda in the UK. Our evaluation found that there was a concern from the academy that the impact agenda may begin to undermine 'blue skies' research and encourage focus towards more applied questions. There was a concern expressed by participants in our evaluation that the impact agenda may move research towards areas that can more easily demonstrate impact and away from areas where impact is less easily demonstrated. This shift could happen at a sector and an individual researcher level. However, it is important to note that REF, as an assessment exercise of all research, recognises and rewards both academic and wider societal excellence (currently weighted at 65% through research outputs compared to 20% for the impact element), and therefore it could be argued that it is the responsibility of the senior management within HEI to provide guidance and an incentive structure to promote a balance within their department or institution.

The representativeness of impacts captured through the exercise Since the publication of the case studies submitted, a number of studies have showcased the great impact of research in a particular area or from a particular funder, such as international development (Hinrichs-Krapels et al. 2015), or the National Institutes for Health Research (NIHR) (Kamenetzky et al. 2016; Morgan Jones et al. 2016). Whilst the case studies themselves are useful anecdotal examples of the achievements of academic research across society, the assessment process necessarily requires a selection process by the sector of what they perceive to be the best examples, and, more importantly in some cases, which examples best adhered to 'the rules'. Therefore there were concerns that the impact case studies submitted may not be representative of the actual impact of research in the sector. As one academic interviewee put it, 'It is a sliver of what impact actually is going on. There is still a lot of other impact work that we do which wasn't included' (Manville et al. 2015a, p. 24). Over time it is important to ensure that these other types of activities, if felt to be valuable, continue to be conducted and mechanisms are in place to ensure they are rewarded appropriately.

5 Discussion: What have we learned in relation to the impact agenda?

Our evaluations, and wider work we have done or drawn on in this area, suggest several areas for consideration and discussion if we are to think about the role of impact in the assessment of research and the implications for the sector going forward. While the previous section has addressed the first question concerning this paper, did the process work, this discussion focusses more on the second main question of this paper: what kinds of impacts did we see?

The first point is that the assessment of impact as part of REF 2014 was the first time anywhere in the world that the impact of nationally funded research was assessed at such a scale. 6975 case studies were produced, demonstrating that universities were able to identify and articulate their impacts. Moreover, our evaluation of the panel assessment showed that panellists felt it was possible to assess and differentiate between different kinds of impacts presented in the case studies. The peer review process was a good way to do this and the case studies were an appropriate format to present a wide range of research. In other words, the process worked.

Following from this, the second point is that while the case studies are useful, they are only snapshots; selected windows into the diverse and multifaceted ways in which research can and is having an impact on society in the UK. However, in many ways they are only representative of one part of a much larger picture. We know from our evaluations that many kinds of impacts were not put forward because they did not fit the rules. There are several potential, and not mutually exclusive, reasons for this: the research wasn't underpinned by sufficiently high quality research, the right kind of supporting evidence was impossible to obtain, the route to impact was complex, the attribution and/or contribution of them to the eventual impact is not easily summarised, or there simply hadn't been enough time for the full impacts to be realised. This doesn't mean the exercise failed, it simply means we must be wary of its limitations, whilst acknowledging what it did achieve.

Moreover, the case studies cannot be aggregated or 'added up' in any neat way due to the lack of standardisation and metrics across them. In other words, not every case study has quantifiable metrics of a similar nature which could be aggregated to present a single picture of the impacts of any given sector. Each case study is unique, each story complex in its own way. This is the nature of research and the reflexive and iterative interactions it has with both its producers and its users. The study from Kings College London and Digital Science also stressed this point, demonstrating the diversity of research that coalesced to create different combinations of impact (KCL and Digital Science 2015). Their study identified 3709 unique pathways to impact across the case studies submitted.

This leads us to a third issue, which is developing our understanding of the kinds of impact categories which will be most appropriate for assessment. This includes consideration of 'impact metrology' and whether impact 'indicators' are appropriate and is an overarching challenge of assessing and evidencing research impact. The evidence we have gathered suggests even if development of a set of impact indicators were possible, they would have to be diverse and multi-faceted in order to work across disciplines. This was one of the main strengths of the case study based approach to impact in REF 2014. For all the limitations of the case studies mentioned above, they did allow for flexibility in understanding and presenting a wide range of impacts. Future analyses are, and should continue to, explore this diversity and celebrate it. Standardising impact will inevitably lead to a narrowing of the productive contributions research can make to society. Just as a

one size approach will not fit all disciplines, a diverse array of methods and indicators is needed to measure and assess impact.

Finally, in other work, the authors have argued that the increased interest in measuring research impact is grounded in a number of different drivers. These drivers can be described as the four ‘A’s’ of research assessment: advocacy, accountability, analysis and allocation. Each driver has a slightly different rationale for it, and when used as a lens through which to view any given research impact exercise, present slightly different perspectives on how impact might be assessed, evidenced and presented (see Morgan Jones and Grant 2013). We would argue that any further consideration of the role of research impact evaluation in future needs to carefully consider the nature of the drivers behind the assessment, and choose methods and approaches accordingly. Explicit acknowledgement of these drivers can have important implications not only for the outcomes of the exercises, but can help to effectively guide the methodological development of the field. An example of this is the fact that the REF has a primary driver of allocation: it is an outcome-based, retrospective assessment process aiming to analyse the quality of research, and now impact, that has occurred as a result of academic research conducted in the UK higher education sector. Because of this focus on outcomes, though, it is less concerned with the *process* of how that impact is created. As pointed out above, much of the research contributing to the impacts relies on at least two or more disciplines coming together. Impact, then, seems often to be the result of multi- or inter-disciplinary research. An assessment of this process of impact creation, though, would likely need a different kind of assessment process, and one which may have more of an analysis driver behind it so as to allow for a deeper understanding of how impact occurs. This kind of analysis would allow policy-makers, researchers and funders alike to better support the kind of research which may have a wider impact on society. At the moment it is primarily researchers who are undertaking this kind of analysis (see below), but more can and should be done in this space.

Thus, we would conclude that, overall, the assessment of research impact in the UK is still a work in progress. HEFCE has released the majority of the case studies online and there has been work to systematically review and evaluate them (KCL and Digital Science 2015; Bangar et al. 2015). Various efforts are underway across the research and innovation system to make use of the case studies, including mining them for data,¹³ using them to analyse impacts within disciplines (Hinrichs-Krapels et al. 2015), by funders (Kamenetsky et al. 2015; Digital Science 2015) and for learning about how to achieve impact. This latter point about learning should not be overlooked. Learning was a crucial part of the impact process, one which came across strongly in our evaluations as both a benefit, but also one of the main challenges of both the impact case study preparations and the assessment of impact itself. This learning, above all else, should be highly valued and not let go to waste by dropping the process or significantly altering the system. But, whether this learning will lead to more, or different kinds of impacts on society, is still to be judged. In order to capture the real benefits from the system, we must think carefully about the interplay of the drivers above. An impact assessment process merely implemented for the purpose of allocation, but without thought for how to make the most of the analysis of its labours, may only lead to superficial efforts to meet the requirements, rather than real added value for the

¹³ For example the regional science and innovation audits being conducted in 2016 designed to map out local research, innovation and infrastructure strengths across the UK and help identify and build on the potential of every region across the country by better evidencing investment decisions and highlighting opportunities for businesses to tap into.

system and society. It seems to us that one measure of success of the exercise, then, will be the extent to which these efforts add value (and indeed what kind of value is yet to be determined) to policymaking, research and analysis, of which only time will tell.

6 Conclusion: Where do we go from here? Impact in a changing research landscape

The research system is changing and impact is just one component of that. For example, the way both public and private sectors support and fund research is shifting. There may be less money to invest in research and governments and researchers alike are thinking about how to do more, with less. The private sector, too, is working collaboratively more and more, using public private partnerships and corporate venture capital to support and R&D at all stages of translation.

As a result of exercises like the REF, the way we understand the impacts of research and innovation on the wider system is changing. The 'science of science' is becoming a more clearly defined field, and assessments of the contribution of programmes or areas of research funding to science, technology and innovation have begun to develop solid methodological approaches to inform our understanding of the economic 'payback' and future funding strategies (Guthrie et al. 2016). Recent studies have also determined more precisely the impact of levels of research funding on research performance, as well as wider system-level spillovers (Sussex et al. 2016). This suggests we may be moving closer to identifying ways in which to define impact-like indicators, but there will inevitably always be shortcomings of applying a narrow set of indicators across a wider field. We have seen this in the area of R&D statistics, where we know that metrics like patents and GDP figures do not capture many of the wider benefits of innovation, and in particular do not capture 'hidden innovation' (NESTA 2009). With greater understanding of how research can have an impact, comes a responsibility to appropriately capture the myriad ways in which this can happen.

In the face of these wider, systemic changes, there is perhaps, a larger and more strategic policy question which we have not yet addressed. If governments need information about how the research system benefits society, then what is the best process for doing that? Is it to incentivise researchers to devote more time and resource to thinking about and ensuring their research does have an impact on society, or is it to develop better methods and metrics for how they measure and understand the contributions of research at a wider system level? Does it involve bridging the gap between research and implementation and jointly funding initiatives which go beyond outputs and impacts and directly address outcomes?

Each question raises a series of possible solutions that will require a more coordinated and joined up approach across policymakers, funders, institutions and research users. For example, asking researchers to spend more time and resource on *translating* their findings into impact may mean that there is less time spent on research. This will need to be accompanied by a change in incentive structures not only within higher education institutions, but also within funding agencies and the associated norms of peer review. This is particularly the case if the aim is not only to indirectly incentivise impact and outcome with measurement approaches and tools but also for policymakers and funders of research to actively encourage it. If direct promotion is the direction of travel, a substantial increase in funding of interdisciplinary and implementation research which gets much closer to

understanding the enablers and barriers to adoption and use of research and innovation will be needed. In this situation, the agenda potentially becomes a more radical attempt at systematic change which could see new alignment of incentives, funding and behaviours, including measurement of success and university level policies (in hiring and promotion for example).

To some extent these changes are underway in a piecemeal fashion but desire for construction of a very new academic terrain is unclear and would be complex for multiple reasons. For example, we already know that one of the pitfalls of exercises like the REF is that applied research is not always seen as being of the same quality as basic research. Equally, we saw from the case study analysis that over 78% of the research underpinning impact case studies was from 2 or more disciplines. This suggests interdisciplinary research is an important (though not always necessary) ingredient in achieving impact. But we also know that funding interdisciplinary research presents challenges. There have been a number of initiatives over time at national and global levels to focus on this, as well as attempts to analyse what we really mean by interdisciplinary research (Bromham et al. 2016). A greater focus on impact will mean these sorts of dynamics which relate to the value given to different types of research will need to be addressed head on. This would require a determination across the community of policymakers, funders and academics to travel in profoundly different directions.

Regardless of the path chosen, if this is the strategic direction policy makers would like to go in, then there will inevitably need to be a corresponding shift with regards to the role different kinds of research plays in society. Not all research will immediately, nor ever, have an impact of the kind the REF case studies highlighted. This is perfectly acceptable of course and should be acknowledged, but equally we need better ways of understanding how to measure these contributions and why it is important.

This in itself raises an important question for the future—do impact evaluations at scale (like the REF) need to be done on a regular basis? While the dual-support system in the UK means that there is the need to evaluate the performance of the research sector on a quinquennial basis, it does not necessarily follow that the impact of that research needs to be assessed as well. One of the main drivers of the inclusion of impact in the REF was to incentivise universities to make, and demonstrate, a greater contribution to society. In a time of austerity, the UK government wanted to show how its investments in the research base were providing value for money. But that has now been done. We know that there is a time lag, in many cases significant, between research and wider impact. This begs the question of how much new information about the impact of research in the sector are we going to learn in four years' time, and what will its new value be? Large-scale research impact assessment exercises are valuable and can be used to demonstrate the benefits society gains from supporting research. But we must continue to assess and learn from those that are being done on a national level in order to determine the ways in which they can positively benefit the sector, including contributing to a more productive research system, rather than creating additional burdens.

Acknowledgements The two evaluations described in this study were funded by the Higher Education Funding Council for England (HEFCE).

Compliance with ethical standards

Conflict of interest The authors declare they have no other conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Academy of Medical Sciences (AMS). (2006). *UK Evaluation Forum Medical Research: Assessing the benefits to society*. London: Academy of Medical Sciences. <https://acmedsci.ac.uk/policy/policy-projects/medical-research-assessing-the-benefits-to-society>. Accessed April 10, 2017.
- Bangar, S., Hargreaves, S., & Mountain, G. (2015). Analysis of ageing research in the Research Excellence Framework impact case studies—Value to BSG members: A conference organized by the British Society for Gerontology University of Sheffield.
- Boaz, A., Fitzpatrick, S., & Shaw, B. (2009). Assessing the impact of research on policy: A literature review. *Science and Public Policy*, 36(4), 255–270.
- Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature*, 534(7609), 684–687.
- Bush, V. (1945). *Science: The endless frontier*. Washington DC: United States Government Printing Office.
- Buxton, M., & Hanney, S. (1996). How can payback from health services research be assessed? *Journal of Health Services Research*, 1, 35–43.
- Buxton, M., Hanney, S., & Jones, T. (2004). Estimating the economic value to societies of the impact of health research: A critical review. *Bulletin of the World Health Organization*, 82, 733–739.
- Cabinet Office. (1993). *Realising our potential: A strategy for science, engineering and technology*. London: HM Stationary Office. <https://www.gov.uk/government/publications/realising-our-potential-a-strategy-for-science-engineering-and-technology>. Accessed May 12, 2017.
- Department for Business, Energy and Industrial Strategy (BEIS). (2016). *Building on success and learning from experience an independent review of the Research Excellence Framework*. London: HM Stationary Office. <https://www.gov.uk/government/publications/research-excellence-framework-review>. Accessed April 10, 2017.
- Department of Education, Science and Training. (2006). *Research quality framework assessing the quality and impact of research in Australia: The recommended RQF*. Canberra: Department of Education, Science and Training. https://research.vu.edu.au/ordsite/management/Recommended_RQF_Dec2006.pdf. Accessed May 12, 2017.
- Digital Science. (2015). *REF 2014 impact case studies and the BBSRC*. Swindon: Digital Science.
- DTZ Consulting and Research. (2006). *Research Councils UK: Analysis of the external costs of peer review*. London: RCUK. <http://www.rcuk.ac.uk/documents/documents/prdtz-pdf/>. Accessed May 12, 2017.
- European Science Foundation (ESF). (2012). *The challenges of impact assessment: Working group 2: Impact assessment ESF member organisation forum on evaluation of publicly funded research*. <http://archives.esf.org/coordinating-research/mo-fora/evaluation-of-publicly-funded-research.html>. Accessed April 10, 2017.
- Freeman, C. (2008). *Systems of innovation: Selected essays in evolutionary economics*. Cheltenham: Edward Elgar.
- Glover, M., Buxton, M., Hanney, S., Pollott, A., & Grant, J. (2014). Estimating the returns to UK publicly funded cancer-related research in terms of the net value of improved health outcomes. *BMC Medicine*, 12, 99.
- Grant, J., Brutscher, P-B., Kirk, S. E., Butler, L., & Wooding, S. (2010). *Capturing research impacts: A review of international practice*. A report prepared for the Higher Education Funding Council for England. RAND Europe. https://www.rand.org/pubs/documented_briefings/DB578.html. Accessed May 12, 2017.
- Greenhalgh, T., Raftery, J., Hanney, S., & Glover, M. (2016). Research impact: a narrative review. *BMC Medicine*, 14, 1.
- Guthrie, S., Kirtley, A., Garrod, B., Pollitt, A., Grant J., & Wooding, S. (2016) A 'DECISIVE' approach to research funding: Lessons from three Retrosight studies. Santa Monica, CA: RAND Corporation. http://www.rand.org/pubs/research_reports/RR1132.html. Accessed May 12, 2017.
- Guthrie, S., Wamae, W., Diepeveen, S., Wooding, S., & Grant, J. (2013). *Measuring research: A guide to research evaluation frameworks and tools*. Santa Monica, CA: RAND Corporation. <http://www.rand.org/pubs/monographs/MG1217.html>. Accessed 12 May 2017.

- Hanney, S. R., Castle-Clarke, S., Grant, J., Guthrie, S., Henshall, C., Mestre-Ferrandiz, J., et al. (2015). How long does biomedical research take? Studying the time taken between biomedical and health research and its translation into products, policy, and practice. *Health Research Policy and Systems*, 13, 1.
- Health Economics Research Group (HERG), Office of Health Economics (OHE), RAND Europe (RE). (2008). *Medical research what's it worth? Estimating the economic benefits from medical research in the UK*. London: UK Evaluation Forum.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261.
- Higher Education Funding Council for England (HEFCE). (2011a). *Assessment framework and guidance on submissions (REF 02.2011)*. HEFCE: London. <http://www.ref.ac.uk/pubs/2011-02/>. Accessed May 11, 2017.
- Higher Education Funding Council for England (HEFCE). (2011b). *Decisions on assessing research impact*. London: HEFCE. http://www.ref.ac.uk/media/ref/content/pub/decisionsonassessingresearchimpact/01_11.pdf. Accessed May 17, 2017.
- Higher Education Funding Council for England (HEFCE). (2012). *Panel criteria and working methods (REF 01.2012)*. London: HEFCE. <http://www.ref.ac.uk/pubs/2012-01/>. Accessed May 12, 2017.
- Higher Education Funding Council for England (HEFCE). (2015). *REF 2014 Panel overview reports*. London: HEFCE. <http://www.ref.ac.uk/panels/paneloverviewreports/>. Accessed May 12, 2017.
- Hill, S. (2016). Assessing (for) impact: Future assessment of the societal impact of research. *Palgrave Communications*. doi:10.1057/palcomms.2016.73.
- Hinrichs-Krapels, S., Kamenetzky, A., Borjes, L., & Grant, J. (2015). *The non-academic impact of international development research in UK Higher Education*. London: The Policy Institute at King's. <http://www.kcl.ac.uk/sspp/policy-institute/publications/non-academic-impact-of-international-development-research.pdf>. Accessed May 12, 2017.
- HM Treasury. (2004). *Science and investment framework 2004–2014*. London: HM Stationary Office. http://webarchive.nationalarchives.gov.uk/+/http://www.hm-treasury.gov.uk/spending_sr04_science.htm. Accessed April 10, 2017.
- Illinois Institute of Technology. (1968). *Technology in retrospect and critical events in science (the TRACES study)*. Washington, DC: National Science Foundation.
- Kamenetzky, A., Hinrichs-Krapels, S., Wooding, S., & Grant, J. (2016). An analysis of the impact of research supported by the UK National Institute of Health Research. In Impacts of agricultural research—an approach of societal values: An International Conference organized by the French National Institute for Agricultural Research (INRA). London: The Policy Institute at King's.
- King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, 13(5), 261–276. doi:10.1177/016555158701300501.
- King's College London and Digital Science. (2015). *The nature, scale and beneficiaries of research impact: An initial analysis of Research Excellence Framework (REF) 2014 impact case studies*. Bristol: HEFCE. <http://www.kcl.ac.uk/sspp/policy-institute/publications/Analysis-of-REF-impact.pdf>. Accessed May 12, 2017.
- Kline, S. J., & Rosenberg, N. (1986). An overview of innovation. In R. Landau & N. Rosenberg (Eds.), *The positive sum strategy: Harnessing technology for economic growth*. Washington, DC: DC National Academies Press.
- Lundvall, B.-A. (Ed.). (1992). *National systems of innovation: Towards a theory of innovation and interactive learning*. London: Pinter.
- Manville, C. Morgan Jones, M., Frearson, M., Castle-Clarke, S., Henham, M., Gunashekar, S., et al. (2015a). *Preparing impact submissions for REF 2014: An evaluation: Findings and observations*. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR727.html. Accessed May 12, 2017.
- Manville, C. Morgan Jones, M., Henham, M., Castle-Clarke, S., Frearson, M., Gunashekar, S., et al. (2015b). *Preparing impact submissions for REF 2014: An evaluation: Approach and evidence*. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR726.html. Accessed May 12, 2017.
- Manville, C., Guthrie, S., Henham, M., Garrod, B., Sousa, S., Kirtley, A., et al. (2015c). *Assessing impact submissions for REF2014: An evaluation*. Santa Monica, CA, RAND Corporation. https://www.rand.org/pubs/research_reports/RR1032.html. Accessed May 12, 2017.
- Martin, B., & Irvine, J. (1983). 'Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12(2), 61–90.
- May, R. M. (1997). The scientific wealth of nations. *Science*, 275, 793.
- Molas-Gallart, J., & Tang, P. (2011). Tracing “productive interactions” to identify social impacts: An example from the social sciences. *Research Evaluation*, 20(3), 219–226.

- Morgan Jones, M., Castle-Clarke, S., Manville, C., Gunashekar, S., & Grant, J. (2013). *Assessing research impact: An international review of the excellence in innovation for Australia trial*. Santa Monica, CA: RAND Corporation. http://www.rand.org/pubs/research_reports/RR278.html. Accessed May 12, 2017.
- Morgan Jones, M., & Grant, J. (2013). Making the grade: Methodologies for assessing and evidencing research impact. In *7 essays on impact* (pp. 25–43). DESCRIBE Project Report for JISC. Exeter: University of Exeter.
- Morgan Jones, M., Kamenetzky, A., Manville, C., Ghiga, I., MacLure, C., Harte, E., et al. (2016). *The National Institute for Health Research at 10 years: An impact synthesis: 100 impact case studies*. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR1574.html. Accessed May 12, 2017.
- NESTA. (2009). *The innovation index: Measuring the UK's investment in innovation and its effects*. London: NESTA. https://www.nesta.org.uk/sites/default/files/innovation_index_2009.pdf. Accessed May 12, 2017.
- Office for National Statistics (ONS). (2017). *UK gross domestic expenditure on research and development, 2015*. <https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/researchanddevelopmentexpenditure>. Accessed April 10, 2017.
- Ovseiko, P., Oancea, A., & Buchan, A. M. (2012). Assessing research impact in academic clinical medicine: A study using Research Excellence Framework pilot impact indicators. *Health Services Research*, 12, 478.
- Penfield, T., Baker, M. J., Scoble, R., & Wykes, M. C. (2014). Assessment, evaluations, and definitions of research impact: A review. *Research Evaluation*, 23(1), 21–32.
- Priem, J., Piwowar, H. A., & Hemminger, B. (2012). *Altmetrics in the wild: Using social media to explore scholarly impact*. arXiv <http://arxiv.org/abs/1203.4745>.
- Research Councils UK (RCUK). (2012). *Impact report 2012*. Swindon: Research Councils UK (RCUK). <http://www.rcuk.ac.uk/publications/reports/impact2012/>. Accessed March 26, 2017.
- Research Councils UK (RCUK). (2014). *Pathways to impact*. Swindon: RCUK. <http://www.rcuk.ac.uk/innovation/impacts/>. Accessed May 11, 2017.
- Slote Morris, Z., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding time lags in medical research. *Journal of the Royal Society of Medicine*, 104(12), 510–520.
- Spaapen, J., & van Drooge, L. (2011). Introducing “productive interactions” in social impact assessment. *Research Evaluation*, 20(3), 211–218.
- Spaapen, J., van Drooge, L., Propp, T., van der Meulen, B., Shinn, T., Marcovich, A., et al. (2013). *Social impact assessment methods for research and funding instruments through the study of productive interactions between science and society*. Brussels: European Framework Programme 7. http://www.siampi.eu/Content/SIAMPLI_Final%20report.pdf. Accessed April 10, 2017.
- Sussex, J., Feng, Y., Mestre-Ferrandiz, J., Pistollato, M., Hafner, M., Burrridge, P., et al. (2016). Quantifying the economic impact of government and charity funding of medical research on private research and development funding in the United Kingdom. *BMC Medicine*, 14(32), 1–23.
- Technopolis. (2010). *REF research impact pilot exercise lessons-learned project: Feedback on pilot submissions: Final report*. Technopolis, November 2010. <http://www.ref.ac.uk/background/pilot/>. Accessed April 10, 2017.
- Walter, I., Nutley, S., Percy-Smith, J., McNeish, D., Frost, S. (2004). Improving the use of research in social care practice. SCIE Knowledge Review 07.
- Weiss, C. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426–431.